Pheromone Protein Er-23 Structure and Stabilization in *Escherichia coli*

A Senior Honors Thesis

Presented in Partial Fulfillment of the Requirements for graduation
with research distinction in the undergraduate colleges
of The Ohio State University

By

Calvin Andrew Rhoads

The Ohio State University

May 2016

Project Advisor:

Professor Thomas J. Magliery
Departments of Chemistry and Biochemistry

Thesis Committee:

Dr. Thomas Magliery
Dr. James Cowan
Dr. Michael Ibba

Abstract

Euplotes raikovi is a unicellular protozoan that has developed a complex network of protein pheromones used in self-non-self recognition among mating types of the species. Members of the Er-pheromone family are small, between 38-51 amino acids, disulfide-rich proteins with a characteristic up-down-up three-helix bundle. Er-23 is the largest member of the family with 51 amino acids, 10 of which are cysteine that participate in the formation of 5 individual disulfide bonds. The amino acid sequences participating in the up-down-up helix architecture have been computationally predicted to form beta-sheet secondary structure.

This project seeks to use Er-23 as a model for investigating the relationship between protein structure, stability, and amino acid sequence—an important topic in the design of medically useful proteins. The Anfinsen hypothesis holds that amino acid sequence determines native protein conformation, driven by thermodynamic forces. However, Luporini and Wüthrich hypothesize that the disulfide bonds of Er-23 maintain the helix architecture despite computational prediction of exclusive beta-sheet structure presence.

Here we describe the heterologous-expression of Er-23 in *Escherichia coli* and its characterization using a variety of methods, including the development of an NMR-based protein concentration estimation technique. We intend to solve a highly constrained NMR solution structure of heterologous Er-23 for comparison with the homologous Er-23 structure previously solved using natural abundance. Additionally, we describe the investigation of the role of individual disulfide bonds in the structure

and stability of heterologously-expressed Er-23 through the replacement of each cysteine disulfide pair with both Ala-Val and Val-Ala mutations. Analysis of Cys-free mutant structure and stability will be completed by comparison of wild-type Er-23 using both CD and NMR studies. We also intend to study disulfide importance by performing the famous Anfinsen reversible denaturation experiment using heterologous Er-23.

Preliminary results from CD and NMR studies suggest that Er-23 is adopting the wild-type fold when expressed in the cytoplasm of *E. coli.* Expression and purification of Cys-free mutants has been partially completed with successful results.

Dedication

To the restless spirit of adventure: for happiness is found in new experiences, and not in the monotonous comforts of home.

"The appetite of youth! What a pity it's wasted on young men" -- **Michael Arlen**

Acknowledgments

Dr. Thomas Magliery
Dr. James Cowan
Dr. Michael Ibba
Graduate Student Sidharth Mohan
Graduate Student David Bowles

Funding:

Undergraduate Research Scholarship, College of Arts and Sciences
Undergraduate Education Summer Research Fellowship

Vita

2012 .....Highland High School, Medina, Ohio

2014-2016.....Undergraduate Researcher, Department of Chemistry, The Ohio State University

Fields of Study

Major Field: Microbiology

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1: Protein Structure

An understanding of protein structure is a critical tool in the biochemist's toolbox. A protein, in its most simple form, is a string of amino acids connected by peptide bonds. Knowing the primary structure of a protein is not particularly useful when investigating the role of a protein in a biological system. It is only when the protein begins to fold and assume its secondary and tertiary structure that it becomes biologically active. Disturbances in the native or active structure, resulting in some degree of denaturation, can cause a protein to lose its function[2].

Christian Anfinsen illustrated the biological tendency for a protein to fold into its native structure. Anfinsen studied ribonuclease, a protein that contains eight cysteine residues that form four disulfide bonds in the native structure. He concluded that the driving force in the correct pairing of the four-disulfide bonds was the interactions of side-chain functional groups along the primary structure[3,4]. Through this work, Anfinsen was the first to prove that amino acid sequence determines conformation and conformation determines biological activity[5,6].

Knowing how a protein folds into its tertiary structure is an important step in understanding how structure meets function. Cyrus Levinthal was one of the original biochemists to suggest that a protein follows a certain folding pathway to reach its native structure. He hypothesized that local interactions and condensations of portions of a folding protein followed a well-defined sequence of events in nature[7].

Levinthal later developed what is known as the Levinthal Paradox, which states that the folding interactions of a protein cannot be random[2]. Mathematical analysis of a relatively small protein consisting of 101 amino acids reveals an astounding $5x10^{47}$ possible unique conformations. Even given an incredible sampling rate, random searches for a correct tertiary structure could take many years. Zwanzig showed that by imposing an energetic bias for a correct conformation, the folding process is mathematically shortened to a biologically-relevant time scale[8].

The protein folding sequence follows a path to reach a most thermodynamically favorable conformation. This is not to say that the native tertiary structure is the most favorable/ lowest energy state. It is the lowest free-energy state among conformations not confined behind high free-energy barriers[9,10].

The combined efforts of Anfinsen and Lenvinthal can summarized as the hypothesis of thermodynamic control of protein folding[10]. The contributions of these biochemistry pioneers have allowed for a more clear understanding of the process of protein folding. However, the prediction of the tertiary structure of a protein from the primary structure remains elusive.

1.2: Disulfide Bonds

Disulfide bonds in proteins are formed by the reversible oxidative condensation of thiol groups in two cysteine residues. The result is a covalent bridge between cysteine residues close in space but not necessarily in primary structure. Disulfide formation is typically sequestered in an oxidative environment favorable to the reaction. This corresponds to the endoplasmic reticulum in eukaryotes and the

periplasm in prokaryotes. Additionally, protein disulfide isomerases (PDI) and disulfide bond proteins (Dsb proteins) assist in the disulfide formation[11].

Reducing environments, such as the prokaryotic cytoplasm, oppose the formation of disulfide bonds. The periplasm of *Escherichia coli* has a reducing potential much lower than its cytoplasm, which helps explain the need to isolate disulfide formation to the periplasm[12]. The formation of stable disulfides in the cytoplasm is an extremely rare event in nature[13]. Proteins with structural disulfides are confined to non-cytosolic intracellular compartments or are located extracellularly[13].

The difficulty in forming disulfide bonds in prokaryotic organisms is a not futile effort. While an oxidized thiol pair is not particularly beneficial within the cytoplasm, it can be advantageous for extracellular proteins. In some extracellular proteins, disulfide bonds are important for correct folding or functionality[11,12]. Another benefit disulfides provide is increased stability in a hostile extracellular environment. The introduction of a new covalent bond places a conformational constraint on the protein backbone[14]. This has a negative effect on the entropy of the denatured state, which makes the native state more preferable[11,14]. The thermodynamic effect of the disulfide on the protein increases the protein's stability with respect to high temperatures, extreme pH, and high concentrations of organic solvents[14].

The stability provided by disulfide bonds will subsequently improve proteins' resistance to proteolytic degradation and oxidants encountered extracellularly. The collective result of the stability benefits provided by disulfide bonds is an increased protein half-life[11,14]. Half-life of a secreted protein is critical as it provides a measure of how long the protein provides its beneficial function to the organism.

The benefit effects are not confined only to wild-type proteins containing disulfide bonds. The possibility of engineering a disulfide bond in a protein has been illustrated by studies on a variety of proteins. An engineered disulfide increased the thermostability of pyroglutamyl peptidase I and *Rhizomurcor miehei* lipase over their wild-type counterparts[15,16]. Both the half-life and thermostability of a tetrameric malate dehydrogenase from *Choloroflexus aurantiacus* were increased when a disulfide bond was introduced[17].

The folding of disulfide-containing proteins is a more complicated process that that of a typical protein. Their folding requires the concerted force of non-covalent bonds within the sequence of amino acids channeling the protein folding process with the formation of the covalent disulfide linkage. The term describing this phenomenon is "oxidative folding" and several models have been developed. Some proteins require the preliminary formation of disulfide bonds in the unfolded state to limit the search in conformational space as the protein backbone folds into its correct tertiary structure. The need for folding catalysts such as PDIs to shuffle non-native disulfides in this oxidative folding model has been suggested[11]. This model highlights the importance of disulfides reducing conformational entropy to help guide protein folding. Work on other proteins has demonstrated a different model of oxidative folding. In this model, the conformational folding of the protein backbone into its tertiary structure is followed by the formation of disulfide bonds[11]. This model highlights the importance of the primary structure to guide protein folding and orient the appropriate residues in physical space such that the disulfide bonds can be formed.

1.3: Pheromone Er-23

Pheromone Er-23 is a member of a structurally homologous group of pheromones produced by protozoan *Euplotes raikovi* used for autocrine growth and paracrine mating signals[18,19]. The collection of Er-pheromones is special due to all the members being small (38-15 amino acids), disulfide-rich proteins[1,20]. Er-23 is a deviant member of known Er-pheromones due to its increased size, number of disulfide bonds, and three-dimensional structure. The Er-23 protein is assembled of 51 amino acid residues, and contains five helical structures. Ten of its residues are cysteine, which participate in the formation of five specific disulfide bonds in the native structure[1].

Er-23 contains a three-helix bundle architecture, arranged in an antiparallel up-down-up arrangement, common to all Er-pheromones. The helix bundle forms a highly compacted core where only 10% or less of the surface area of core residues is solvent-accessible[1]. Also common to all Er-pheromones, protein secondary structure prediction of Er-23 suggests the exclusive presence of beta-sheet secondary structure in the segments composing the four alpha helices. It has, thus, been predicted that three of the disulfide bonds in Er-23 stabilize and enforce the characteristic up-down-up helix bundle in the compacted core[1].

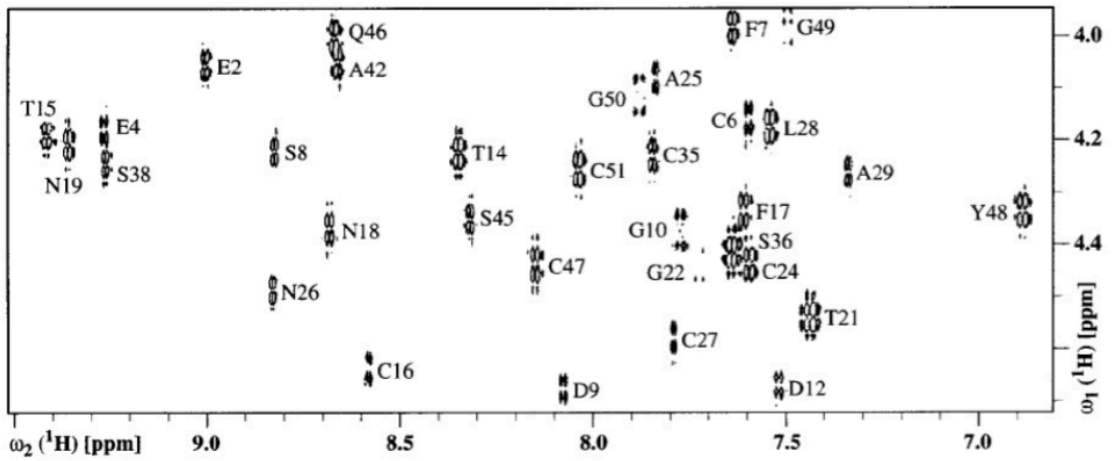

Figure 1.1: [^{1}H,^{1}H]-2QF-COSY NMR spectrum of homologously expressed Er-23[1]

The NMR solution structure has been previously solved for Er-pheromones following purification from the native organism and ^{13}C assignments using natural abundance[1].

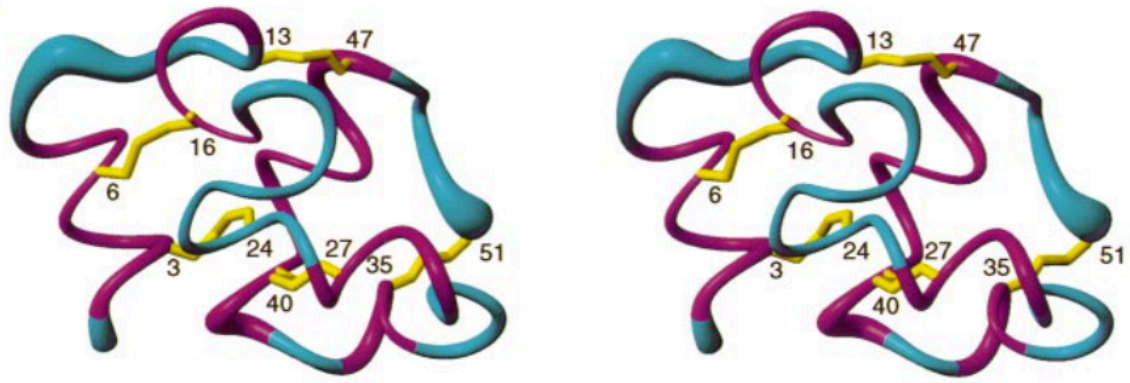

Figure 1.2: Stereo views of Er-23 NMR structure calculation[1]

Temperature-induced denaturation studies on Er-23 using circular dichroism has been completed following purification from *E. raikovi*[21]. It was concluded that WT

Er-23 was the most stable of the Er-pheromones as it was not denatured when heated to 95°C.

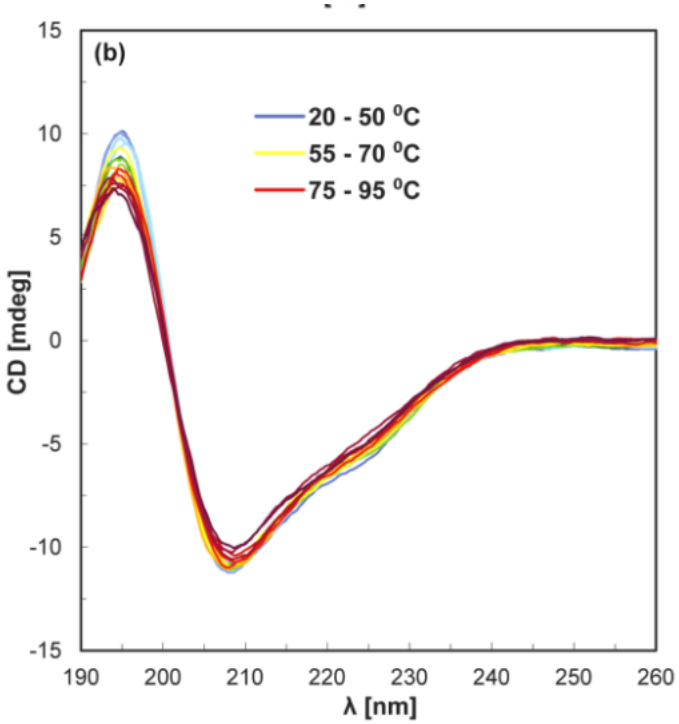

Figure 1.3: Thermal melt CD spectra of homologously expressed Er-23[21]

Previous work in the Magliery lab has prepared the Er-23 gene for expression in *Escherichia coli* and began the characterization via circular dichroism experiments of wild-type Er-23 expressed in *E. coli*[22]. This preliminary work has alluded to the high stability and native folding of Er-23 expressed in *E. coli*. Circular dichroism spectra of Er-23 from *E. coli* show a significant peak at 209 nm, indicating the presence of disulfide bonds and alpha-helical structure.

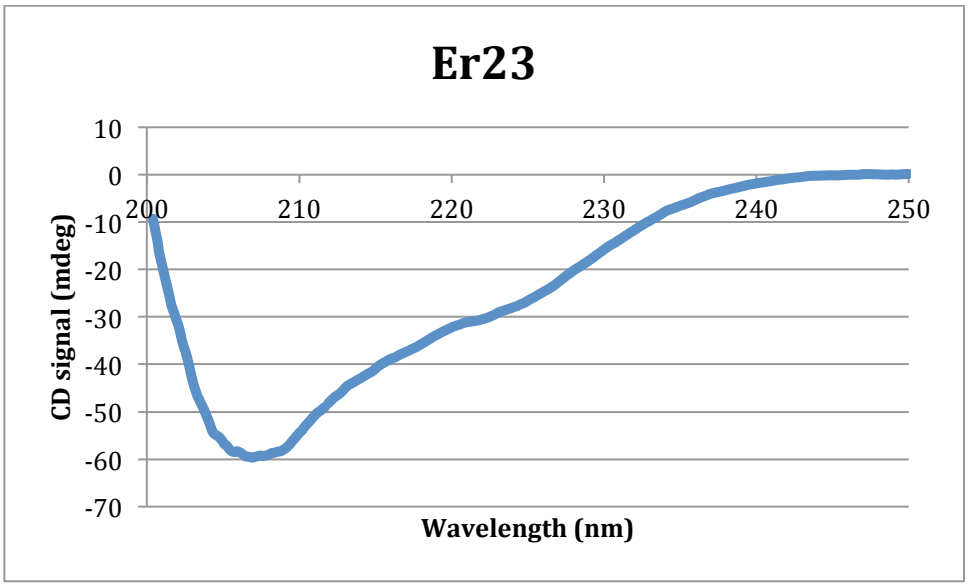

Figure 1.4: CD spectrum of heterologously expressed Er-23[22]

When acquiring thermal melt data, the CD signal at 222 nm remained relatively unchanged then Er-23 was heated to 95°C. CD spectra of Er-23 samples before and after thermal melts showed that the samples subjected to 95°C maintained the spectra of unheated samples[22].

1.4: Expression and Purification of Disulfide-rich proteins in *Escherichia coli*

In the initial stages of the Er-23 project, it was decided that we would express and purify Er-23 from *E. coli* cells, a heterologous organism. This was decided out of convenience rather than practicality. Previous work outside of the Magliery lab expressed and purified Er-23 from its native organism, *Euplotes raikovi*[19,23]. Therefore, an expression strategy had to be designed before characterization of Er-23 could begin. As discussed in section 1.2, disulfide bond-containing proteins are not often found in the cytoplasm of *E. coli*.

The thiol-disulfide redox potential of the cytoplasm of *E. coli* is too low to provide a driving force for the formation of disulfide bonds[13]. The cytoplasm has a reducing redox potential of -240 mV to -270 mV, where as the periplasm is more oxidizing environment at -0.165 mV[24]. Additionally, there are no active enzymes that catalyze thiol-oxidation in the cytoplasm of *E. coli*[13]. However, the Bessette lab developed a fast-growing *E. coli* strain with null mutations at the *trxB* and *gor* loci[13]. These genes are responsible for products that maintain cytoplasmic thioredoxins and glutaredoxins in a reduced state. In the absence of these genes, thioredoxins and glutaredoxins are in an oxidized state and can catalyze the formation of protein disulfide bonds. The *trxB* and *gor* mutant has a cytoplasmic thiol-disulfide redox potential suitable for disulfide formation[13,24].

The cytolasmic thiol-disulfide redox potential of the double mutant created by Bessette was intermediate that of the typical periplasm and cytoplasm, which allows for slow disulfide bond formation kinetics in comparison to in the periplasm. The slower oxidation rate is favorable for disulfide-rich proteins because it the disulfide formation is more heavily influenced by conformational preferences of the amino acid sequence[13].

1.5: Synthesis of Cysteine-free proteins

The removal cysteine residues is a topic of interest in the Magliery lab. Cysteine residues can complicate the folding because the formation of disulfide bonds can be a slow step. Additionally, cysteine residues can be problematic for the storage of proteins in vitro because the oxidation of cysteine residues can lead to misfolding or oligimerization instead of the formation of correct disulfide bonds[25]. Lastly, cysteine removal can improve the protein's expression in *E. coli*[26]. Previous work in cysteine

removal in the Magliery lab has focused on constructing a Cys-free variant of Rop, a well-characterized protein.

After deciding to remove cysteine residues, the question then becomes which amino acids should be substituted for cysteine. The strategy is to find residues that could replace the cysteine pair and best conserve the structure of the native protein. When constructing the Cys-free Rop variant, it was decided to substitute in a serine pair, alanine pair, valine pair, or alanine/valine pair. The function of Rop in *E. coli* cells is to control the copy number of ColE1 plasmid number. Activity of Rop variants can be assessed using a GFP screen that links the expression of green fluorescent protein with the ColE1 plasmid copy number[27]. The activities of the Cys-free Rop variants were compared to wild-type Rop in the GFP screen to determine the variants that most closely represented wild-type[25].

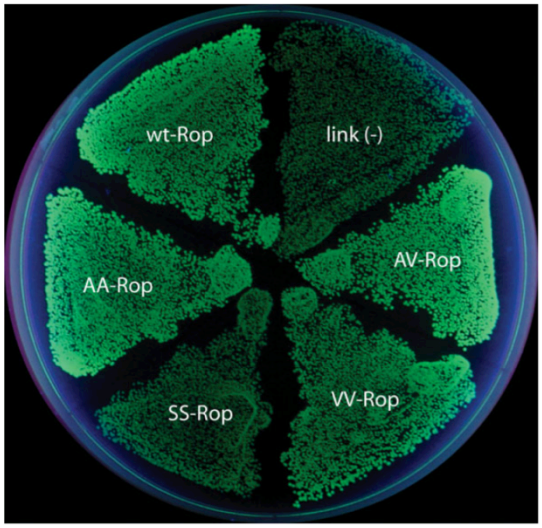

Figure 1.5: In vivo activity screen of Cys-free Rop variants[25]

Despite both the Ala/Ala and Val/Val variants being active, only the Ala/Ala variant could be expressed at significant levels using the T7 promoter[25]. The screen

showed that the Ala/Val variant had activity most similar to wild-type, and had similar helical structure as determined by CD. Interestingly, the Ala/Val variant had increased thermal stability over wild-type.

The findings of Cys-free Rop study formed the basis for disulfide removal in Er-23. Previous work in the Magliery lab has begun to replace Cys/Cys pairs of Er-23 disulfide bonds with Ala/Val or Val/Ala pairs to determine which, if any, of the disulfide bonds are needed for the correct folding and stability of Er-23[22]. These findings will allow us to determine the level of contribution of disulfides in the structure and properties of Er-23.

Chapter 2

Objectives

2.1: Synthesis and Characterization of WT Er-23

The primary goal of this project is to successfully express WT Er-23 pheromone in *E. coli*. This work has been previously begun by Venuka Durani[28] and later continued by Kiersten Lessig[22]. Characterization will continue using an Er-23-pHLIC plasmid (created by Venuka Durani[28]) maintained in DH10B DE3 *E. coli* and expression in Origami DE3 *E. coli*. Following optimized expression and purification of WT Er-23, characterization via NMR, crystallography, and CD thermal melts will be performed.

2.2: Synthesis and characterization of Er-23 Cys-free variants

Our secondary goal of this project is to study the implications of disulfide bond removal in Er-23. The first step of synthesizing cys-free Er-23 variants involves the replacement of cysteine residues participating in individual disulfide bonds with Ala/Val or Val/Ala pairs. This work has been partially completed by Kiersten Lessig of the Magliery Lab[22]. We will then work towards successful expression and purification of these variants similar to WT Er-23. Characterization of the successfully purified cys-free variants via CD thermal melts and other studies will be performed to analyzed variant stability and structure in comparison to WT Er-23.

Chapter 3

Materials & Methods

3.1: Synthesis of WT Er-23 and Cys-free variants

Gene Construction

Methods of overlap PCR were used for assembly of gene fragments, and the reactions were performed using an Eppendorf thermal cycler. The WT Er-23 genes were synthesized from four primers (sequences shown in Table 3.1). Variant genes were synthesized from four primers containing the specific mutations, which replaced a specific disulfide cysteine residue pair with alanine and valine residues (primers shown in Table 3.2). The genes were assembled from two separate overlap PCR reactions. First, a Klenow reaction was used to synthesize the body of the gene from the two respective primers. During the second reaction, forward and reverse amplification primers were used to amplify the Klenow product. Amplification reactions were carried out in a total volume of 100 μL with an annealing temperature of 60 °C for 25 cycles using Pfu polymerase.

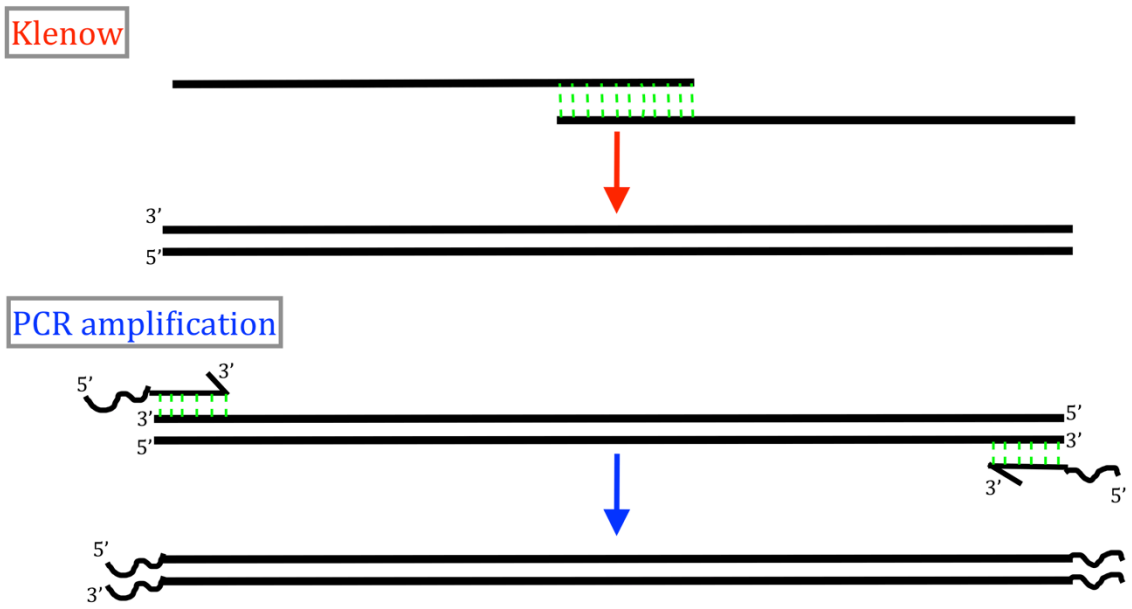

Figure 3.1: Overlap PCR cloning schematic

WT Er-23	Construct	GGCGAATGCGAACAGTGCTTTAGCGATGGCGGCGATTGCACGACCTG CTTCAATAATGGCACCGGCCCATGCGCGAATTGCCTGGCGGGCTATCC GGCGGGGTGCAGCAATAGCGACTGTACCGCGTTTCTGAGCCAGTGCT ATGGAGGTTGC
	Amp_Fwd	AATATTCCATGGCGAAAATTTATATTTCCAGGGCGAATGCGAACAGT GC
	Amp_Rev	ATTATAAGGATCCGTTTAGGACTTCCAATGTTAGCAACCTCCATAGC ACTGGC
WT Er-23 GSSG	Klenow_1	GGCGAATGCGAACAGTGCTTTAGCGATGGCGGCGATTGCACGACCTG CTTCAATAATGGCACCGGCCCATGCGCGAATTGCCTGGCGGGCTATCC GGCGGGG
	Klenow_2	GCAACCTCCATAGCACTGGCTCAGAAACGCGGTACAGTCGCTATTGC TGCACCCCGCCGGATAGCCCGCCAGGCAATTCGCGCATGGGCCGGTGC C
	Amp_Fwd	AAAATTTATATTTCCAGGGTAGTAGTGGCGAATGCGAACAGTGC
	Amp_Rev	TTATCCACTTCCAATGTTAGCAACCTCCATAGCACTG

Table 3.1: Primers used to synthesize WT Er-23 constructs

Er-23 13A47V	Klenow_Fwd	GGCGAATGCGAACAGTGCTTTAGCGATGGCGGCGATGCGACAACATG TTTTAACAACGGGACCGGTCCTTGCGCGAACTGTCTTGCTGGG
	Klenow_Rev	GCAACCACCATACACTTGACTCAGAAAAGCAGTACAGTCGCTATTAGA ACATCCAGCCGGGTACCCAGCAAGACAGTTCGCGCAAGGACC
	Amp_Fwd	CATCATGGCCATGGCGAAAATTTATATTTACAAGGCGAATGCGAACA GTGCTTT
	Amp_Rev	AATAATAATCCGGATCCTTAGCAACCACCATACACTTG
Er-23 13V47A	Klenow_Fwd	GGCGAATGCGAACAGTGCTTTAGCGATGGCGGCGATGTGACCACCTGC TTTAACAACGGTACCGGTCCTTGCGCGAACTGTCTTGCTGGC
	Klenow_Rev	GCAACCACCATACGCCTGAGACAGAAAAGCAGTACAGTCACTATTACT ACAACCAGCAGGATAGCCAGCAAGACAGTTCGCGCAAGGACC
	Amp_Fwd	CATCATGGCCATGGCGAAAATTTATATTTACAAGGCGAATGCGAACA GTGCTTT
	Amp_Rev	AATAATAATCCGGATCCTTAGCAACCACCATACGCCTG
Er-23 27A40V	Klenow_Fwd	GGCGAATGCGAACAGTGCTTTAGCGATGGCGGCGACTGTACCACCTGC TTTAACAACGGTACTGGTCCTTGTGCAAATGCTCTAGCTGGG
	Klenow_Rev	GCAGCCACCATAGCATTGAGACAGAAACGCAGTCACATCAGAGTTGCT ACAACCAGCCGGATACCCAGCTAGAGCATTTGCACAAGGACC
	Amp_Fwd	CATCATGGCCATGGCGAAAATTTATATTTACAAGGCGAATGCGAACA GTGCTTT
	Amp_Rev	AATAATAATCCGGATCCTTAGCAGCCACCATAGCATTG
Er-23 27V40A	Klenow_Fwd	GGCGAATGCGAACAGTGCTTTAGCGATGGCGGCGATTGTACCACCTGT TTTAACAACGGTACCGGTCCTTGTGCGAACGTGCTGGCTGGG
	Klenow_Rev	GCAGCCACCATAGCACTGAGACAGAAACGCAGTCGCATCAGAGTTAGA ACAGCCAGCAGGATACCCAGCCAGCACGTTCGCACAAGGACC
	Amp_Fwd	CATCATGGCCATGGCGAAAATTTATATTTACAAGGCGAATGCGAACA GTGCTTT
	Amp_Rev	AATAATAATCCGGATCCTTAGCAGCCACCATAGCACTG
Er-23 5	Klenow_Fwd	GGCGAATGCGAACAGTGCTTTAGCGATGGCGGCGATTGCACCACCTGC

		TTTAACAACGGTACTGGACCATGTGCGAACTGCCTG
	Klenow_Rev	CACGCCGCCATAGCACTGACTCAGAAACGCGGTGCAATCGCTGTTGCT CGCTCCCGCTGGATATCCTGCCAGGCAGTTCGCACATGGTCC
	Amp_Fwd	ATTATTCCATGGCGAAAATTTATATTTACAGGGCGAATGCGAACAGT GC
	Amp_Rev	AATAATAATGGATCCTTACACGCCGCCATAGCACTG
Er-23 35V51A	Klenow_Fwd	Same as Er-23 35A51V Klenow_Fwd
	Klenow_Rev	CGCGCCGCCATAGCACTGACTCAGAAACGCGGTGCAATCGCTGTTGCT CACTCCCGCTGGATATCCTGCCAGGCAGTTCGCACATGGTCC
	Amp_Fwd	Same as Er-23 35A51V Amp_Fwd
	Amp_Rev	AATAATAATGGATCCTTACGCGCCGCCATAGCACTG

Table 3.2: Primers used to synthesize Er-23 Cys-free variants

Cloning

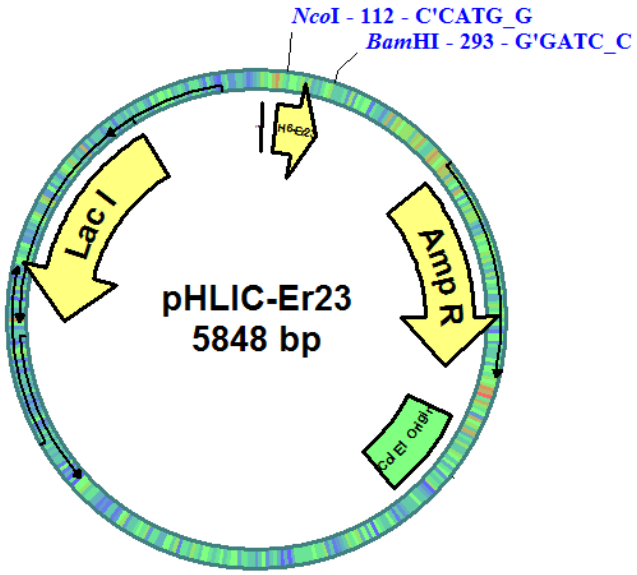

Figure 3.2: pHLIC vector with Er-23 gene

Figure 3.3: Er-23 protein design

Following the overlap PCR steps, each gene insert made from its respective four primers contained an Nco*I* and Bam*HI* restriction site. The gene inserts were then digested at 37 °C for 4 hours with Nco*I* and Bam*HI* restriction endonucleases. These inserts were then ligated with pHLIC vector (constructed by Venuka Durani, Brandon Sullivan, and Thomas Magliery), which was also digested at its Nco*I* and Bam*HI* restriction sites at 37 °C for 4 hours in a total volume of 50 µL. The ligation reaction of the digested inserts and pHLIC vector occurred overnight at 16 °C in a total volume of 5 µL. The vector complete with a gene insert was then transformed via electroporation into electrocompetent DH10B *E. coli* cells. The transformed cells were plated on LB agar containing 100 µg/mL ampicillin and incubated at 37 °C overnight. The resulting isolated colonies were picked and their DNA was extracted using Qiagen miniprep protocol. The DNA was sequenced via Sanger sequencing by Genewiz.

3.2: Expression and Purification of WT Er-23 and Cys-free variants

Expression

Liquid media used for expression was 2YT with 100 µg/mL ampicillin, unless otherwise stated. WT Er-23 and Cys-free variant DNA was transformed into electrocompetant Origami™ B *E. coli* cells: F^- *ompT hsdS*$_B$(r_B^- m_B^-) *gal dcm lacY1 ahpC*

(DE3) *gor522*:: Tn*10 trxB* (Kan^R, Tet^R) from EMD Millipore. The *trxB* and *gor* mutations are selectable on kanamycin and tetracycline, respectively[29]. These cells were plated on LB agar containing 100 μg/mL ampicillin and incubated at 37 °C overnight. Small 25 mL seed cultures of isolated colonies were grown overnight. The seed cultures were pitched into 2 L of media in 4 L baffled flasks, which were grown at 37 °C in a shaker incubator until reaching OD_{600}=0.8, about 4 hours. The cultures were then induced with 0.1 mM IPTG during the log growth phase. Expression occurred overnight at 25 °C. Following expression, the cells were centrifuged and cell pellet frozen at -80 °C until purification.

Protein Purification

The frozen cell pellet was resuspended in 15 mL of lysis buffer (50mM Tris-HCl, 300 mM NaCl, 10 mM imidazole, pH 8). To the resuspended cells, 5 mM $MgCl_2$, 0.5 mM $CaCl_2$, 5 μL of both DNase I and RNase A (each at 5 mg/ml), and 0.1% Triton X-100 were added. The mixture was then incubated at 4 °C or on ice for at least 30 minutes. Following the incubation, the cells were lysed via an Emulsiflex at a pressure differential between 15,000 and 20,000 psi. The cell suspension mixture passed through the Emulsiflex at least three times and remained on ice between cycles.

The homogenized cell lysate was then centrifuged at 17,000 rpm for 45 minutes. The supernatant was allowed to bind to 1.5 mL Ni-NTA agarose resin (Qiagen) at 4 °C for 60 minutes on a nutator. The mixture was poured through a large (20 mL) pre-fitted column, and the flow-through collected. The column was then washed with 12 mL wash buffer (50mM Tris-HCl, 300 mM NaCl, 20 mM imidazole, pH 8) and 12 mL wash buffer 2

(50mM Tris-HCl, 300 mM NaCl, 50 mM imidazole, pH 8). Lastly, the column was eluted with 3 mL elution buffer (50mM Tris-HCl, 300 mM NaCl, 250 mM imidazole, pH 8).

To the collected elution sample was added 0.08 mg TEV protease twice; first, for 60 minutes at room temperature on the nutator and next, overnight under the same conditions. The sample was then dialyzed back into lysis buffer using a 2 kDa MWCO Slide-A-Lyzer™ (Thermofisher) dialysis cassette. The sample was dialyzed twice, each for 60 minutes in 1 L lysis buffer at room temperature.

The dialyzed sample was allowed to bind to 1 mL Ni-NTA agarose resin for 60 minutes at 4 °C on the nutator. After binding, the sample was poured through a small (10 mL) pre-fitted column, and the flow-through collected and saved as the cleaved protein. The column was washed with 12 mL lysis buffer and eluted with 3 mL elution buffer. The cleaved protein could be optionally concentrated by injecting the flow-through into a 2 kDa MWCO Slide-A-Lyzer™ dialysis cassette. The surface of the cassette membrane was coated with Aquacide II (Millipore) and held at room temperature until the desired volume reduction was achieved. Concentrated cleaved protein was removed from dialysis cassette and stored at room temperature for future characterization.

3.3: Characterization of WT Er-23 and Cys-free variants

SDS-PAGE

All protein samples throughout purification were saved. Protein samples were mixed in equal volume of 2x SDS loading buffer without reducing agents. Samples were

loaded, along with a protein ladder, into their respective lanes of an 18% SDS gel. A gel electrophoresis machine (Bio-Rad) was used to run the SDS gel at 200 V until the blue dye band reached the bottom of the gel. The gel was immediately removed and treated with various staining procedures.

Silver Staining

All staining buffers were made fresh prior to use for best results. 18 MΩcm H_2O was used in all staining buffers and washing steps. Gel and all materials used in staining procedure were first washed thoroughly with H_2O to remove impurities. The gel was fixed in fixing solution (50% methanol, 12% acetic acid, H_2O) for 30 minutes. Fixed gel was washed 3x for 10 minutes with 50% methanol. The gel was then pretreated for 1 minute in pretreatment solution (1.3 mM $Na_2S_2O_3$, H_2O), followed by 3x 20 seconds washing with H_2O. Then the gel was nutated in stain solution (12 mM $AgNO_3$, 0.075% formaldehyde, H_2O) for 20 minutes, then washed 2x 30 seconds with H_2O. Developer solution (0.57M Na_2CO_3, 0.05% formaldehyde, 0.025 mM $Na_2S_2O_3$, H_2O) was added to the gel until bands appeared on the gel. The development was stopped by transferring gel to the fixing solution for 10 minutes.

MALDI-TOF MS

MALDI matrix (0.05% TFA, 50% acetonitrile, saturated sinapinic acid) was made fresh immediately prior to use and otherwise stored in the dark at room temperature. MALDI matrix was mixed in equal volume with protein purification samples. No other treatment of protein purification samples was required. The protein-matrix mixture

was spotted on a clean MALDI plate in volumes between 0.5 µL and 1.5 µL. The spotted MALDI plate was allowed to dry completely overnight prior to the MALDI experiment. The experiment was performed using a Bruker microflex MALDI-TOF MS operated by Sidharth Mohan.

Amino Acid Analysis

Cleaved protein flow-through of WT Er-23 GSSG was submitted to Jinny Johnson, M.S. of Texas A&M University for amino acid analysis. The flow-through sample was dialyzed into Er-23 storage buffer (50mM phosphate, 50mM NaCl, pH 6) using a 2 kDa MWCO Slide-A-Lyzer™ dialysis cassette to remove the Tris buffer. The dialyzed sample was concentrated using Aquacide II (Millipore) as described earlier. The concentration was adjusted to 1 mg/mL based on concentration estimation of the cleaved 6xHis tag in the TEV cleaved sample lane on an 18% Coomassie-stained SDS gel.

1D NMR

Cleaved protein flow-through sample of WT Er-23 GSSG (from same purification as used in amino acid analysis) was dialyzed into Er-23 storage buffer using a 2 kDa MWCO Slide-A-Lyzer™ dialysis cassette to remove imidazole and Tris. Then the dialyzed sample was concentrated using Aquacide II (Millipore) as described earlier. The sample was Pasteur pipetted into an NMR tube as a mixture of 90% concentrated cleaved protein flow-through and 10% D_2O. The NMR scans were acquired and analyzed by David Bowles of the Magliery Lab.

3D NMR

The expression and purification was conducted as described earlier, except rich 2YT media was replaced with Neidhart's minimal media (1x Neidhart's solution, 19 mM NH_4Cl, 1.32 mM K_2HPO_4, 10 mM $NaHCO_3$, 50mg/L thiamine, 17 mM glucose, 18 MΩ/cm^2 H_2O). ^{15}N-labeled NH_4Cl was used as the nitrogen source and ^{13}C-labeled glucose was used as the carbon source. To the Neidhart's minimal media was added 100 μg/mL ampicillin and 30 μg/mL Kanamycin (final concentrations) prior to inoculation. The concentrated cleaved protein sample dialyzed into Er-23 storage buffer was Pasteur pipetted into a clean NMR tube with 10% D_2O. The NMR scans were acquired and analyzed by David Bowles of the Magliery Lab.

Chapter 4

Results

4.1: Synthesis and Characterization of WT Er-23

Previous work on the protein outside the Magliery lab chose to express and purify Er-23 directly from *Euplotes raikovi*. Given the unusual demands of culturing *Euplotes raikovi*, it was decided we would use *Escherichia coli*, the standard organism used in the Magliery lab, in the expression of Er-23. This required that we first synthesize the Er-23 gene such that it could be expressed in *E. coli*.

The first challenge in synthesizing WT Er-23 was to decide the expression scheme. Our selected expression scheme had to reflect its unique composition of amino acids, almost one-fifth of them being cysteine residues. All 10 of the cysteine residues participate in the formation of 5 specific disulfide bonds in the native structure. The expression scheme selected must create conditions under which the formation of the disulfide bonds in the native structure would be possible. Typically, proteins with disulfide bonds are expressed in the cell periplasm, where the reduction potential is amiable to thiol oxidation. Additionally, the periplasm contains foldases, DsbA and DsbC, which catalyze the formation and isomerization of disulfide bonds.

On the other hand, cytoplasmic expression is the usual method employed in the Magliery lab. Venuka Durani decided to proceed with a cytoplasmic expression scheme, and synthesized the WT Er-23 gene for that method. The Origami™ B (DE3) *E. coli* strain was selected as the expression cell line to aid in disulfide formation. The WT Er-

23 gene product was for immobilized metal ion affinity chromatography (IMAC) purification. Therefore, the protein product was designed to have an N-terminal 6xHis-tag that would be cleaved off at the adjacent TEV protease site. These features were followed by a GSSG linker, and then the WT Er-23 protein. The GSSG linker was added under the assumption that it would aid in accessibility of the TEV site by TEV protease.

Early attempts at protein purification optimization were fraught with confusion. Coomassie stain is the standard SDS-PAGE protein stain in the Magliery lab. When a typical protein is purified using our protocol, it would appear uncleaved in the pre-TEV elution lane, then separate into distinct cleaved protein and 6xHis-tag bands in the post-TEV lane, and finally stain as purified cleaved protein in the second flow-through lane. However, WT Er-23 formed one distinct band in the post-TEV lane and was entirely unseen in the second flow-through lane.

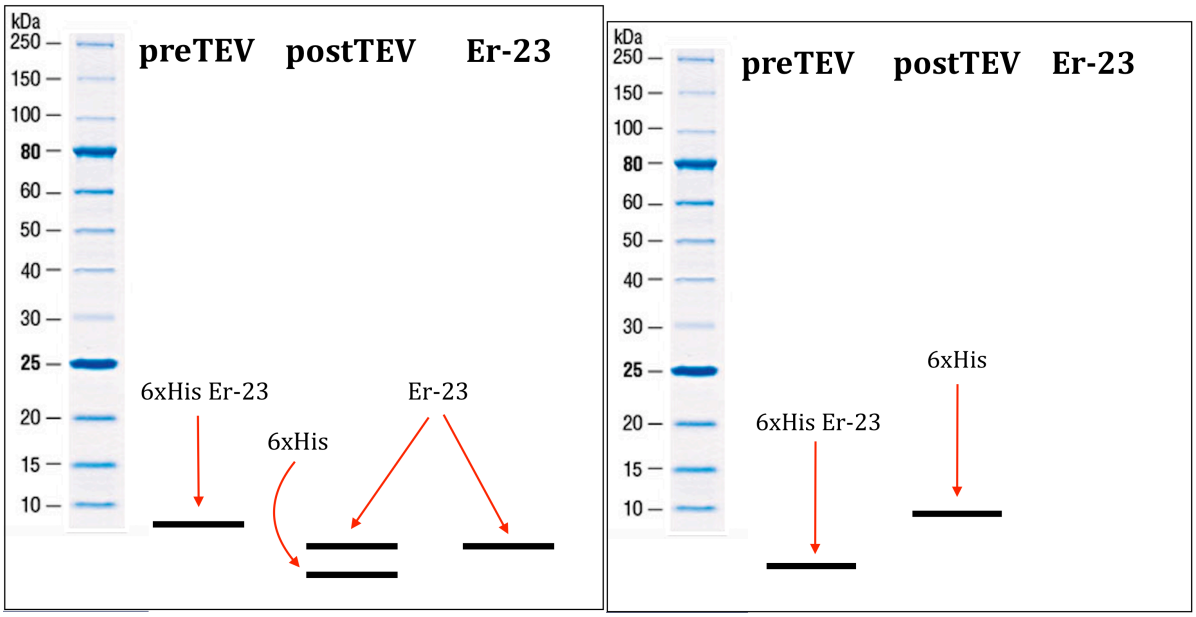

Figure 4.1: Expected vs. Actual Er-23 purification SDS gel

MALDI-TOF mass spectrometry experiments on the WT Er-23 purification samples proved the presence of the various protein products in their respective lanes, based on predicted mass. The mass spectrum below shows uncleaved WT Er-23 GSSG (7393 Da) in the pre-TEV sample, cleaved 6xHis-tag (2075 Da) in the post-TEV sample, and cleaved WT Er-23 GSSG (5336 Da) in the post-TEV and flow-through samples.

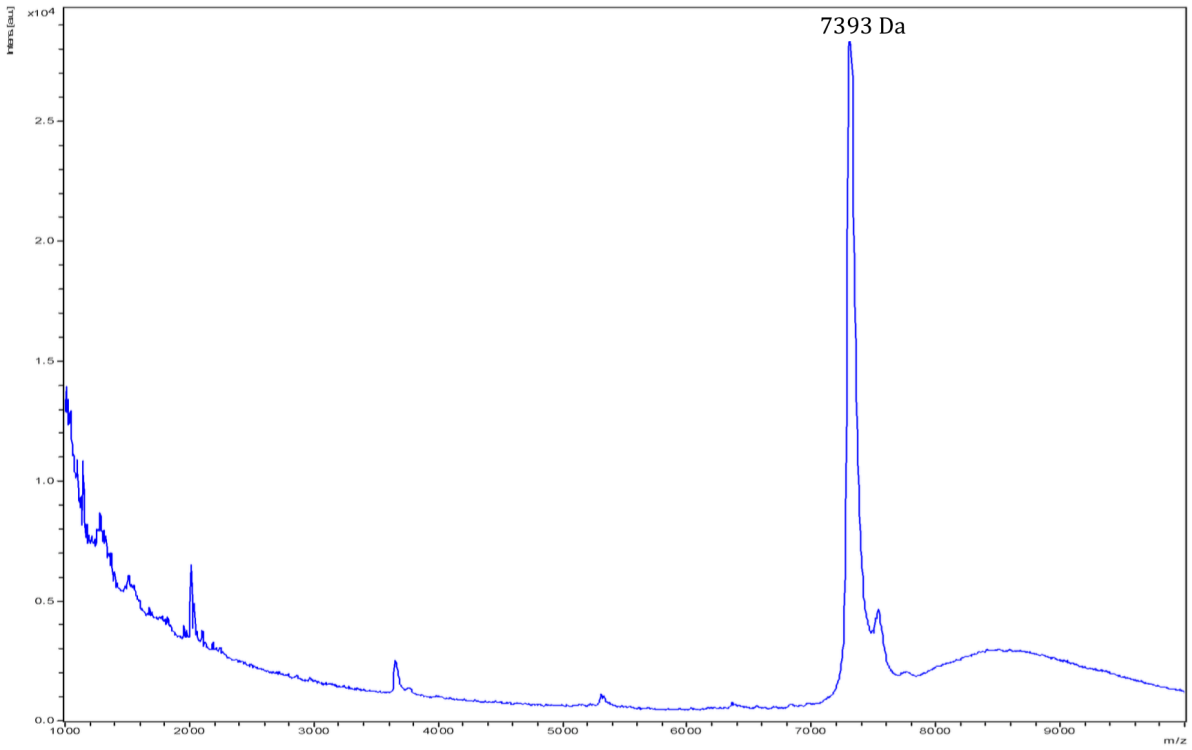

Figure 4.2: MALDI spectra of preTEV reaction 6xHis WT Er-23 GSSG (7393 Da)

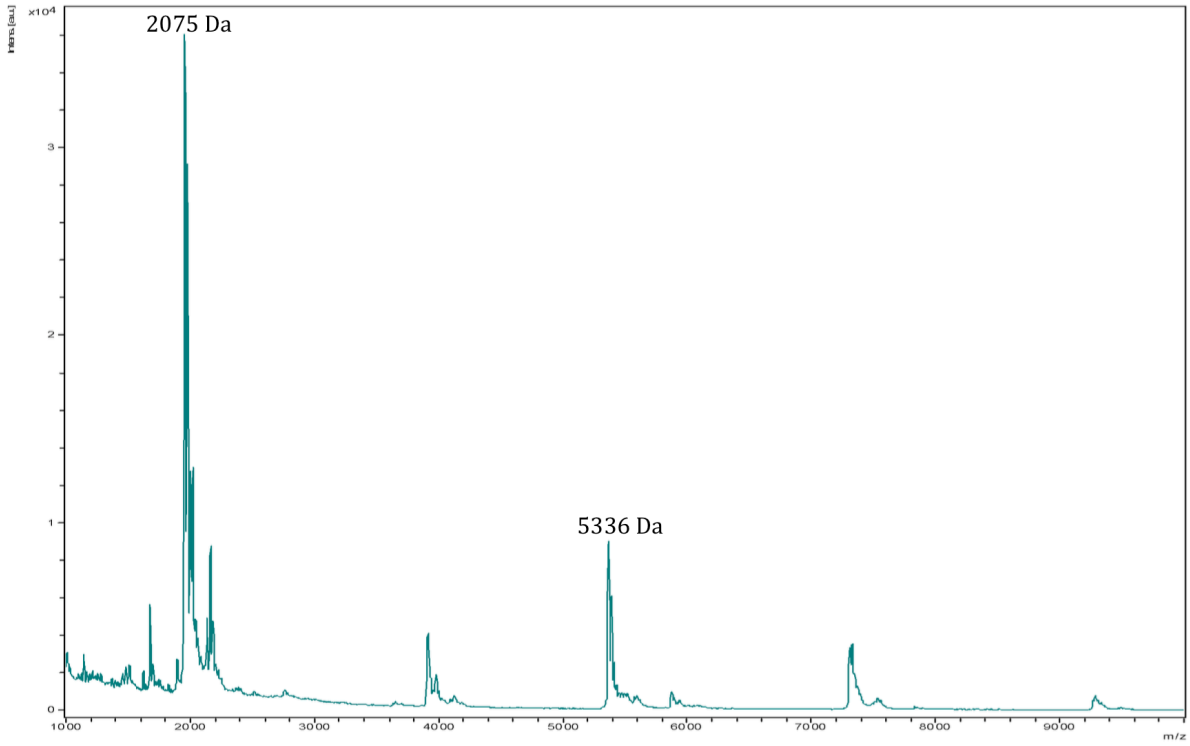

Figure 4.3: MALDI spectra of postTEV reaction 6xHisTag (2075 Da) and GSSG WT Er-23 (5336 Da)

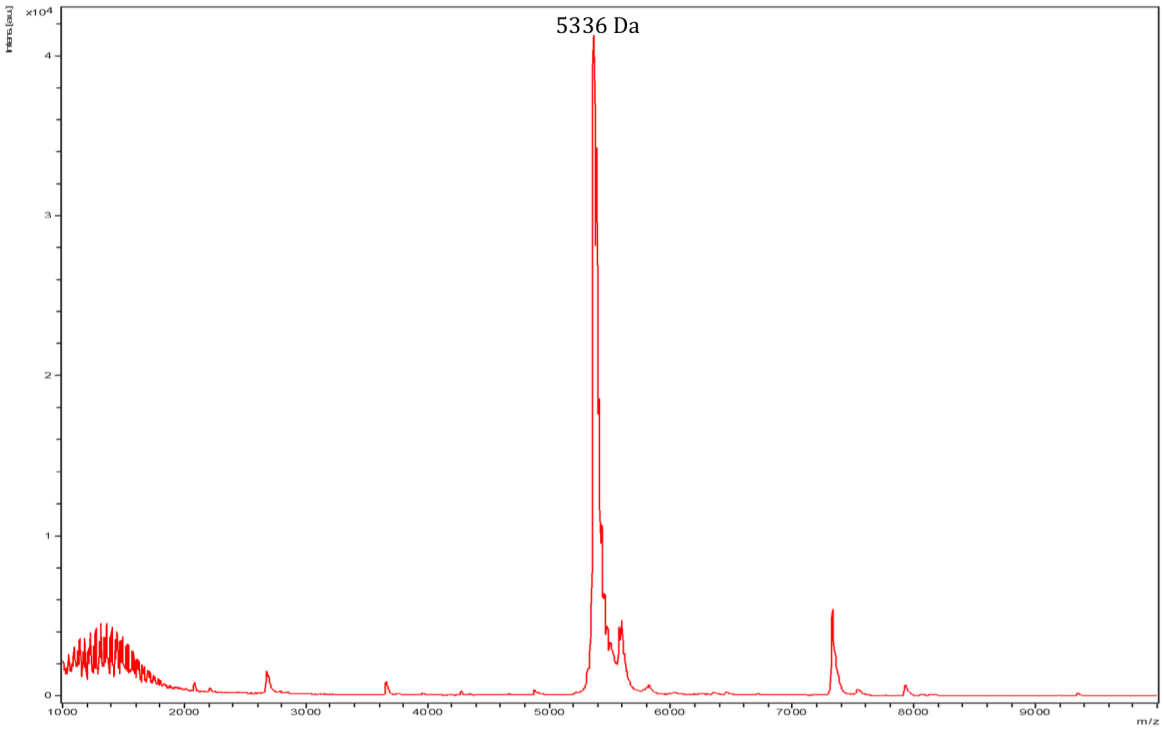

Figure 4.4: MALDI spectrum of the second flow-through WT Er-23 GSSG (5336 Da)

Alternate SDS gel staining procedures were attempted to produce an SDS gel with cleaved WT Er-23 visualized. The typical method of concentration estimation of a protein sample in the Magliery lab requires a quantitative protein staining procedure, such as Coomassie stain. The only successful staining method we found was silver staining, which is not a quantitative stain. The pictures below are replicate SDS gels of WT Er-23 GSSG purification samples. The FT2 labels on the gels correspond to the last flow-through of the protein purification, which contains the purified WT Er-23 GSSG. As should be noted, bands appear near the bottom of the silver stained SDS gel in the post-TEV and FT2 lanes that are not visible when a replicate of the gel was stained with Coomassie. This experiment provided additional verification that WT Er-23 GSSG was purified using our developed protocol, but the WT Er-23 GSSG was not binding to Coomassie stain.

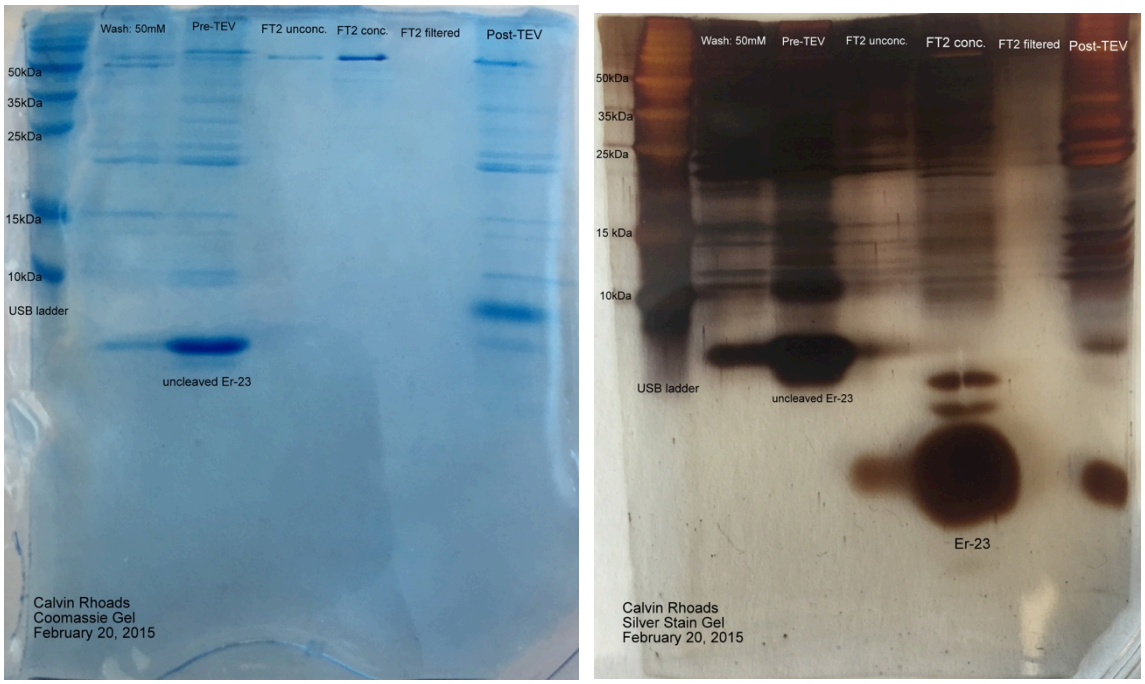

Figure 4.5: Er-23 purification SDS gel stained with Coomassie vs. Silver Stain

The next goal in the project was to find a method to confidently estimate the concentration of Er-23 for concentration-dependent experiments, since Coomassie and other quantitative protein stains were found to be ineffective. We purified a large master stock of WT Er-23 GSSG, and then divided it into three aliquots. One of the aliquots was sent to Texas A&M University for amino acid analysis. This provided us with the estimated concentration of master stock of WT Er-23 GSSG.

From the second aliquot of the master stock a ^{1}H NMR spectrum was acquired. The proton NMR signal is concentration-dependent, so the peak intensity of the acquired spectrum reflects the concentration of WT Er-23 GSSG in the master stock, which was known based on the results of amino acid analysis from the same sample.

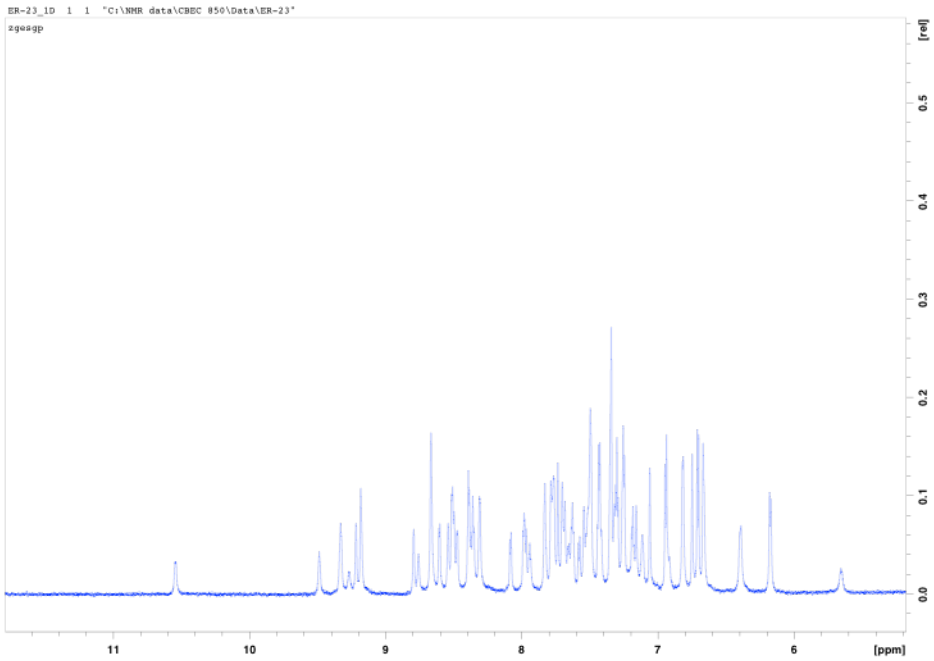

Figure 4.6: WT Er-23 GSSG ^{1}H NMR

Once the purification issues had been resolved, we were able to shift our focus to characterization of Er-23. The major characterization goal of this project was to resolve the structure of Er-23 when expressed in *E. coli*, and compare our findings to the previously resolved structure of Er-23 expressed in *Euplotes raikovi*. After obtaining the proton NMR spectrum, we went on to acquire a ^{1}H-^{15}N HSQC NMR of WT Er-23 GSSG using a ^{15}N-labelled sample. The ^{1}H-^{15}N HSQC spectrum contains 54 well-defined cross-peaks, which were assigned by David Bowles. Other backbone NMR experiments were performed such as HNCO, HN(CA)CO, HNCA, and HN(CO)CA. Side-chain experiments were performed such as an ^{1}H-^{13}C HSQC, CC(CO)NH, and H(CCO)NH. Additionally, NOESY experiments were performed, which will lead us towards solving to solution structure.

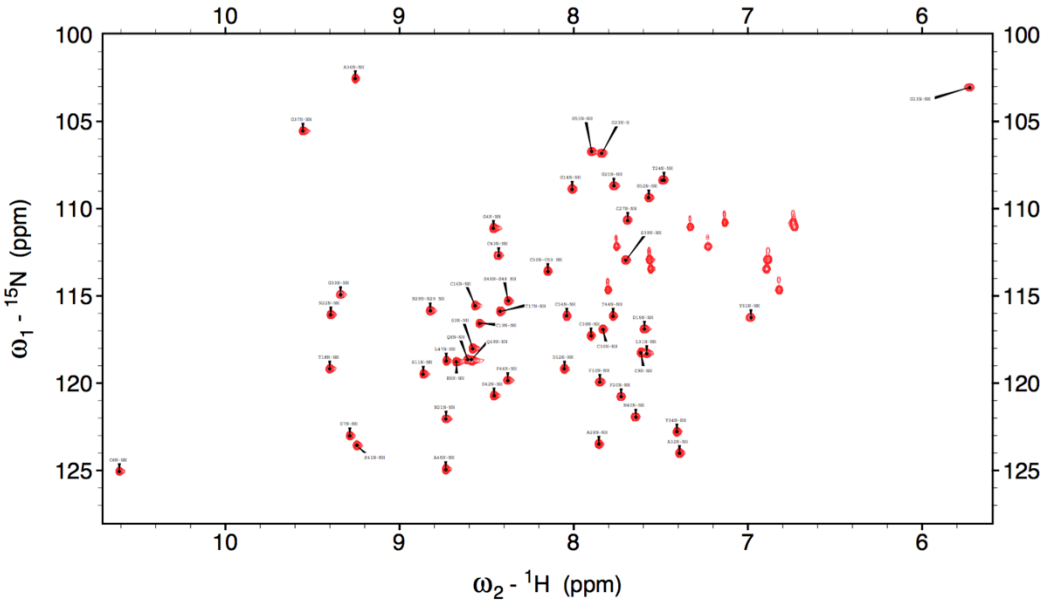

Figure 4.7: ^{1}H-^{15}N HSQC NMR of WT Er-23 GSSG

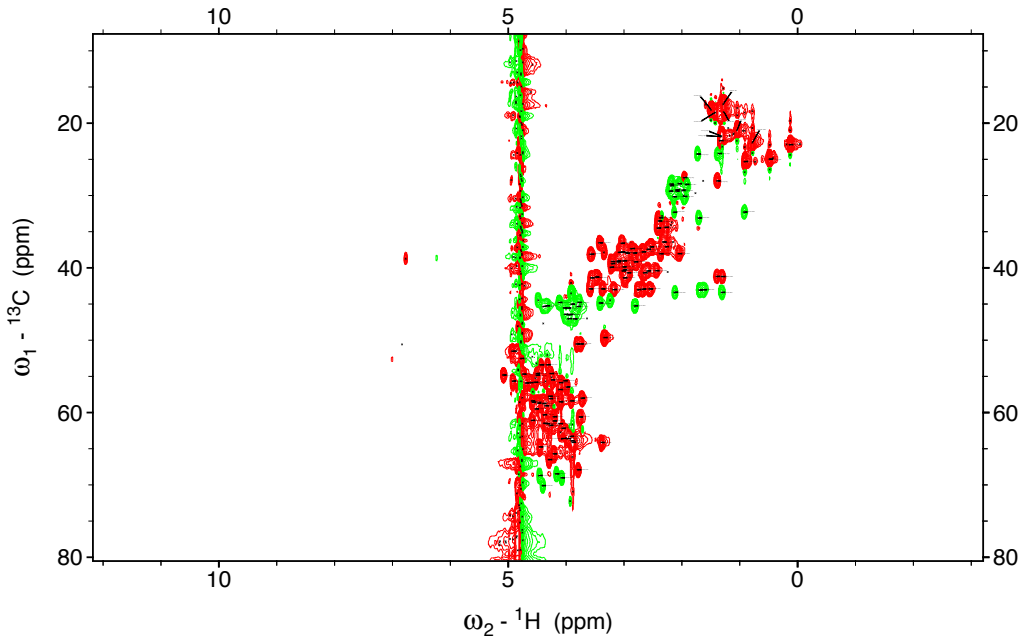

Figure 4.8: ^{1}H-^{13}C HSQC NMR of WT Er-23 GSSG

4.2: Synthesis and Characterization of Cys-free variants

Former undergraduate researcher, Kiersten Lessig, began synthesis of the Cys-free variants. Both variants for the 3&24 and 6&16 disulfide bonds were created. These variants were initially made because computational stability analysis (via FoldX) predicted that those variants would have stability most similar to WT Er-23[22]. These Cys-free variants were designed without the GSSG linker that was used in the wild-type Er-23 construct that had been synthesized and partially characterized[22]. No explanation for the lack of GSSG linker was provided. Regardless of this unresolved issue, we decided to continue synthesizing the Cys-free variants without GSSG linkers.

Cys-free variant synthesis has been completed and the variants have been analyzed on an SDS gel for expression and TEV cleavage. For each variant, the preTEV and postTEV purification fractions are analyzed. We look for the 6xHis-Er-23 band in the preTEV and the 6xHis band running higher on the SDS gel in the postTEV fraction, as is seen with WT Er-23 on SDS gels when stained with Coomassie. Both of the 3&24 variants do not appear to express in the figure below. The 6&16 variants are being expressed, but no shift in banding pattern between pre- and postTEV fractions is observed, indicating that the variants are not cleaved by TEV protease.

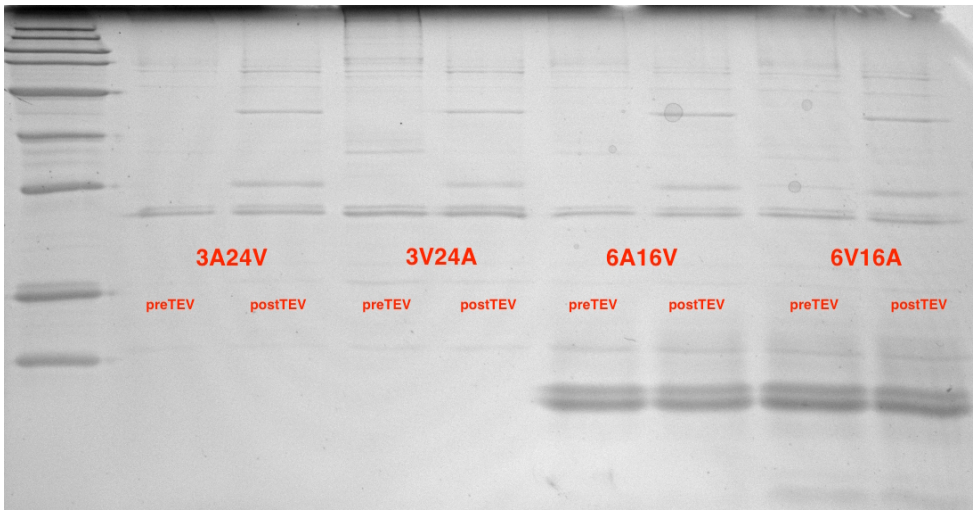

Figure 4.9: 3&24 and 6&16 Er-23 Cys-free variant SDS gel

The 13&47, 27&40, and 35&51 Cys-free variant purification SDS gel is pictured below. All of the variants below have an obvious shift in band size between the pre- and postTEV fractions similar to WT Er-23 purification fractions, which suggests that the variants of these three disulfide bonds are capable of expressing in *E. coli* and cleaving with TEV protease using our protocol.

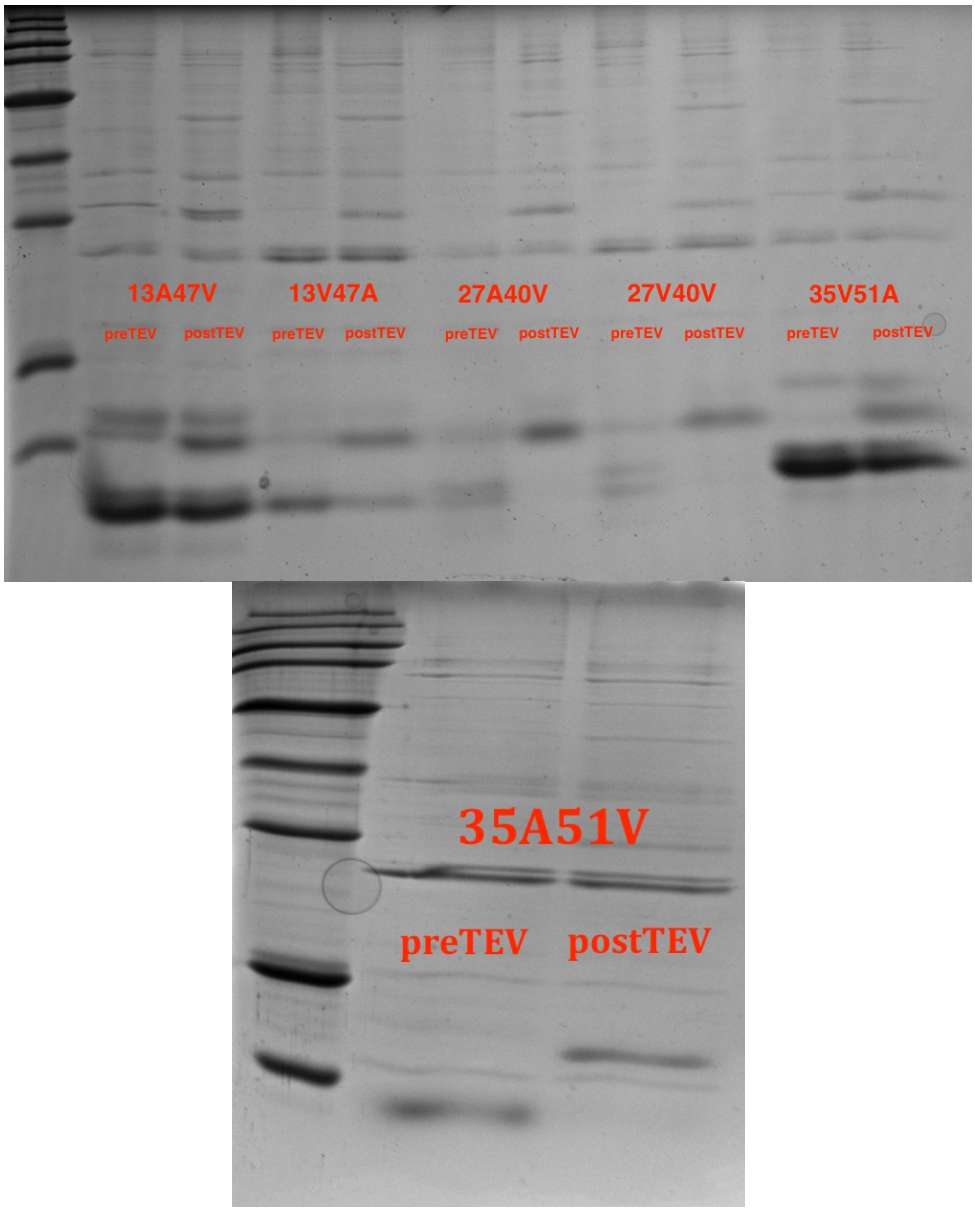

Figure 4.10: 13&47, 27&40, and 35&51 Er-23 Cys-free SDS gel

Chapter 5

Discussion and Future Work

In this project, we have performed a series of experiments that provide evidence suggesting successful heterologous expression of Er-23 in *Escherichia coli*. MALDI Mass Spectroscopy spectra of Er-23 purification fractions indicate that the products of protein expression and purification are present in their expected purification fractions based on predicted size. Circular dichroism studies of WT Er-23 reveal signal in the 210-220 nm range indicates the formation of alpha-helical structure in the protein, and the dampened signal indicates the presence of disulfide bonds. Additionally, the CD spectra of Er-23 from homologous expression is similar to that from *E. coli*, which can be evidence of correct folding of Er-23 in *E. coli*. Furthermore, similarity of thermal melt CD spectra between homologously- and heterologously-expressed Er-23 is evidence that the Er-23 is adopting a wild-type fold and therefore wild-type thermal stability, when expressed in *E. coli*. Finally, NMR studies provide evidence of well-folded structure of heterologously-expressed WT Er-23 due to well defined peaks in ^{1}H NMR and well defined cross peaks in ^{1}H-^{15}N HSQC NMR. The combination of this evidence supports the conclusion that Er-23, a small disulfide-rich protein, is folding into the wild-type conformation when expressed in the cytoplasm of *Escherichia coli*.

The NMR solution structure of heterologously-expressed WT Er-23 GSSG is being calculated during the writing of this thesis. Upon resolution of the solution structure, comparison of heterologous Er-23 structure and homologous Er-23 structure

will confirm whether heterologous Er-23 is adopting wild-type fold when expressed in typically unfavorable conditions. If the protein is adopting the wild-type fold, our experiment gives us the ability to solve a much more constrained NMR solution structure for Er-23 than previously done given the isotopically labeled Er-23 used in our experiments. Additionally, completion of the Anfinsen experiment on WT Er-23 will provide some evidence for what is driving its fold: amino acid sequence or disulfide formation.

The goal of synthesizing and characterizing the Er-23 Cys-free variants is well underway. All ten variants have been synthesized, and five of them are capable of expression and TEV cleavage as determined on an SDS gel. The 3&24 Cys-free variants are unable to be expressed based on our protocol, which may suggest that this disulfide bond is critical to the protein, expression optimization is required, or that the variants provided are in some way flawed. The 6&16 Cys-free variants are capable of expression, but not able to be TEV cleaved. This could suggest that steric occlusion of the TEV site is obstructing the access of TEV protease to its substrate (TEV site). The addition of a GSSG linker in the 6&16 Cys-free construct could improve TEV protease access.

Future work on the Cys-free variants should be focused on purification/expression optimization and characterization. Purification and expression optimization is needed to determine how many of the Cys-free variants are capable of expressing in *E. coli*, which can provide initial evidence for the contribution of each disulfide to the wild-type fold. MALDI spectra of purification fractions of the variants should be obtained to confirm the presence of expected protein products in each

fraction by mass. Finally, Cys-free variant structure analysis should be performed and compared to WT Er-23. CD spectra should be obtained to investigate the formation of secondary structure and disulfide bonds. Thermal melts may also be performed to compare Cys-free thermal stability to WT Er-23. Alternatively, ^{1}H or ^{1}H-^{15}N NMR spectra of the Cys-free variants can be obtained to assess whether they are well-folded. Evidence from the Cys-free variant structural studies will allow us to determine the contribution and importance of each disulfide bond in the fold and stability of WT Er-23. The investigation of disulfide contribution to the fold of Er-23 will allow us to assess whether the fold is determined by the disulfides as previously predicted[1], or by the amino acid sequence as hypothesized by Anfinsen[5,6].

References

1. Zahn, R., Damberger, F., Ortenzi, C., Luporini, P. & Wüthrich, K. NMR structure of the Euplotes raikovi pheromone Er-23 and identification of its five disulfide bonds. *J. Mol. Biol.* **313,** 923–31 (2001).

2. Levinthal, C. How to fold graciously. *Mössbauer Spectrosc. Biol. Syst. Proc.* **24,** 22–24 (1969).

3. Anfinsen, C. B. The Formation and Stabilization of Protein Structure. *Biochem. J.* **128,** 737–749 (1972).

4. Anfinsen, C. B. & Haber, E. ARTICLE : Studies on the Reduction and Re-formation of Protein Disulfide Bonds Studies on the Reduction and Re-formation of Protein Disulfide Bonds. 1361–1363 (1961).

5. Kresge, N., Simoni, R. D. & Hill, L. Work of Christian Anfinsen. (2006).

6. Anfinsen, C. B. Principles that Govern the Folding of Protein Chains. *Science (80-.).* **181,** 223–230 (1973).

7. Levinthal, C. Are there pathways for protein folding? *J. Chim. Phys. Physico-Chimie Biol.* **65,** 44–45 (1968).

8. Zwanzig, R., Szabo, A. & Bagchi, B. Levinthal ' s paradox. **89,** 20–22 (1992).

9. Dill, K. A. Dominant Forces in Protein Folding Ken. **29,** (1990).

10. Durup, J. (Universite P. S. et C. On 'Levinthal paradox' and the theory of protein folding. *J. Mol. Struct.* **424,** 157–169 (1998).

11. Arolas, J. L., Aviles, F. X., Chang, J. Y. & Ventura, S. Folding of small disulfide-rich proteins: clarifying the puzzle. *Trends Biochem. Sci.* **31,** 292–301 (2006).

12. Depuydt, M., Messens, J. & Collet, J.-F. How proteins form disulfide bonds. *Antioxid. Redox Signal.* **15,** 49–66 (2011).

13. Bessette, P. H., Aslund, F., Beckwith, J. & Georgiou, G. Efficient folding of proteins with multiple disulfide bonds in the Escherichia coli cytoplasm. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 13703–13708 (1999).

14. Bulaj, G. Formation of disulfide bonds in proteins and peptides. *Biotechnol. Adv.* **23,** 87–92 (2005).

15. Kabashima, T., Li, Y., Kanada, N., Ito, K. & Yoshimoto, T. Enhancement of the thermal stability of pyroglutamyl peptidase I by introduction of an intersubunit disulfide bond. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **1547,** 214–220 (2001).

16. Han, Z. L., Han, S. Y., Zheng, S. P. & Lin, Y. Enhancing thermostability of a Rhizomucor miehei lipase by engineering a disulfide bond and displaying on the yeast cell surface. *Appl. Microbiol. Biotechnol.* **85,** 117–126 (2009).

17. Bjørk, A., Dalhus, B., Mantzilas, D., Eijsink, V. G. H. & Sirevåg, R. Stabilization of a tetrameric malate dehydrogenase by introduction of a disulfide bridge at the dimer-dimer interface. *J. Mol. Biol.* **334,** 811–821 (2003).

18. Luporini, P., Alimenti, C., Ortenzi, C. & Vallesi, A. Ciliate mating types and their specific protein pheromones. *Acta Protozool.* **44,** 89–101 (2005).

19. Di, G. G. *et al.* A structurally deviant member of the Euplotes raikovi pheromone family: Er-23. *J. Eukaryot. Microbiol.* **49,** 86–92 (2002).

20. Vallesi, A. *et al.* Autocrine, mitogenic pheromone receptor loop of the ciliate Euplotes raikovi: Pheromone-induced receptor internalization. *Eukaryot. Cell* **4,** 1221–1227 (2005).

21. Geralt, M., Alimenti, C., Vallesi, A., Luporini, P. & Wüthrich, K. Thermodynamic stability of psychrophilic and mesophilic pheromones of the protozoan ciliate euplotes. *Biology (Basel).* **2,** 142–50 (2013).

22. Lessig, K. L. Towards the flip of FruR to Er23 via Disulfide Bond Modification. (2014).

23. Concetti, A., Raffioni, S., Miceli, C., Barras, D. & Luporini, P. Purification to Apparent Homogeneity of the Mating Pheromone of mat-1 Homozygous Euplotes raikovP. **261,** 10582–10586 (1986).

24. Gebendorfer, K. & Winter, J. in *Oxidative Fold. Pept. Proteins* 43 (Royal Society of Chemistry, 2009).

25. Hari, S. B., Byeon, C., Lavinder, J. J. & Magliery, T. J. Cysteine-free Rop: A four-helix bundle core mutant has wild-type stability and structure but dramatically different unfolding kinetics. *Protein Sci.* **19,** 670–679 (2010).

26. Lepock, R., Frey, E. & Hallewellj, A. Contribution of Conformational Stability and Reversibility of Unfolding to the Increased Thermostability of Human and Bovine Superoxide Dismutase Mutated at Free Cysteines * C ;;"(T) = AHcg. **265,** 21612–21619 (1990).

27. Magliery, T. J. & Regan, L. A cell-based screen for function of the four-helix bundle protein Rop : a new tool for combinatorial experiments in biophysics. **17,** 77–83 (2004).

28. Durani, V. The Cycle of Protein Engineering: Bioinformatics Design of Two Dimeric Proteins and Computational Design of a Small Globular Domain. *PhD Propos.* **1,** (2012).

29. 70837 | Origami™ B(DE3) Competent Cells - Novagen. *EMD Millipore* (2016).